nature research

Corresponding author(s): Matthew Albert; Etienne Patin

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample size (n=1,000) was chosen so that we have 95% power to detect medium effects of genetic factors, corresponding to 0.6 units of phenotype standard deviation. This assumes a significance threshold of P=10e-10, to account for multiple testing.

2. Data exclusions

Describe any data exclusions.

As to donors of the Milieu Intérieur cohort, exclusion criteria included evidence of, or report a history of neurological or psychiatric disorders, or severe/chronic/recurrent pathological conditions, history or evidence of alcohol abuse, recent use of illicit drugs (including cannabis), recent vaccine administration, and recent use of immune modulatory agents. To avoid the influence of hormonal fluctuations in women during the peri-menopausal phase, only pre- or post-menopausal women were included. To avoid the presence of population structure in our study population (i.e., highly variable genetic backgrounds due to different ancestry), which would impact upon the power to detect genotype-to-phenotype associations, we restricted our study to individuals of European-descent, i.e., French citizens whose ancestry for three generations was of Metropolitan French origin (i.e., the subject's parents and grandparents were born in continental France).

As to flow cytometry data, we removed outlier points as follows: a value in the higher tail of the distributions was considered an outlier if the distance to the closest point in the direction of the mean of the distribution was more than 60% of the total range of the sample, while a value in the lower tail was considered an outlier if that distance was more than 15% of the total range of the sample.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

To assess repeatability of cytometric measures, we analyzed the same sample, in independent runs, by a single operator and run on a single cytometer. In our experimental setting, fresh blood samples from six healthy donors were separated into five aliquots and stained using each of the panels and run on the liquid handling platform, followed by acquisition on the corresponding MACSQuant cytometer. To assess reproducibility of cytometric measures, we evaluated the stability of staining over time, an important consideration for large cohort studies. To provide a stable reference, we utilized commercially available stabilized blood, analyzed over a period of one month.

Candidate variants identified by GWAS were tested for replication in an independent cohort of 75 donors recruited through the Genentech Genotype and Phenotype (gGAP) Registry, California, US.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Blood samples from participants were processed on a daily basis (~8 donors/day, during an entire year). To correct for possible batch effects, the processing day was included as a random effect in all linear models, and all immunophenotypes were corrected for batch effects with ComBat prior to genome-wide association studies. We also verified that intrinsic, environmental and genetic factors that are candidate predictors of immunophenotypes are not impacted by the processing day.

5. Blinding

Describe whether the investigators were blinded to

Investigators who collected samples were blinded to participant demographic

Describe whether the investigators were blinded to variables (age, sex, etc.).

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Flow cytometric data were generated with MACSQuantify™ software version 2.4.1229.1 and FlowJo software version 9.5.3. The impact of non-genetic factors on immunophenotypes was analysed with the with the custom mmi R package (http://github.com/JacobBergstedt/mmi). Genome-wide association studies were performed with GEMMA 0.94.1. The difference in contribution to explained variance between innate and adaptive immunophenotypes was tested using the lme4qtl R package. Other secondary programs can be found in Online Methods.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

Genotypic data will be made available under restriction access on EGA prior to publication. A Data Access Committee (DAC) will verify that academic investigators asking access to the data do not try to re-identify the donors, and that their research complies with the initial aims of the Milieu Intérieur project.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

83 different antibodies were used. Dye, specificity, clone, supplier, catalog and lot numbers of each antibody can be found in Supplementary Table S2. A unique lot of each antibody was used for the entire study. Antibodies were selected based on the following criteria: (i) specificity of the signal, as based on the difference between the positive and the negative populations and the spread of the negative population; (ii) signal resolution; (iii) availability of desired fluorochrome; (iv) fluorochrome stability (tandem dyes); (v) price and availability of single lot of reagents for cohort study; and, when possible, (vi) availability of CE-IVD format.

To validate the assays, we performed experiments to assess repeatability and reproducibility. To assess repeatability, we analyzed the same sample, in independent runs, by a single operator and run on a single cytometer. In our experimental setting, fresh blood samples from six healthy donors were separated into five aliquots and stained using each of the four panels and run on the liquid handling platform, followed by acquisition on the corresponding MACSQuant cytometer. The results were highly repeatable, with intra-panel coefficients of variation below 15% for most of the analyzed cell subsets. To assess reproducibility, we evaluated the stability of staining over time, an important consideration for large cohort studies. To provide a stable reference, we utilized commercially available stabilized blood, analyzed over a period of one month. These data showed reproducible results with CVs in the range of 2.0–5.3%.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

No research animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The 1,000 healthy donors of the Milieu Intérieur cohort were recruited in Rennes, France, and included 500 women and 500 men, and 200 individuals from each decade of life, between 20 and 69 years of age.

nature research

Corresponding author(s): Matthew Albert

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Whole blood samples were collected from the 1,000 healthy, fasting donors on Li-heparin, every working day from 8 to 11AM, from September 2012 to August 2013, in Rennes, France.
6. Identify the instrument used for data collection.	MACSQuant analyzers (Serial numbers 2420 & 2416)
7. Describe the software used to collect and analyze the flow cytometry data.	MACSQuantify™ software version 2.4.1229.1
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Not applicable.
9. Describe the gating strategy used.	Please see Supplementary Figures 1 to 10.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒